Methodological Advisory Committee
June 2002

Using a Generalised Estimation Methodology for ABS Business Surveys

Authors: John Preston and James Chipperfield
c/ABS PO Box 10, Belconnen, ACT 2616
fax: (02) 6252 8015
email: john.preston@abs.gov.au phone: (02) 6252 6970
email: james.chipperfield@abs.gov.au phone: (02) 6252 7301

Part 1: Executive Summary

1. The availability of Business Activity Statement (BAS) data collected by the Australian Taxation Office (ATO) has provided the Australian Bureau of Statistics (ABS) with opportunities to improve the efficiency of sample designs and estimations for its business surveys. The ABS business surveys currently use two methods of estimation; number-raised estimation and ratio estimation. While ratio estimation allows the use of one auxiliary variable to improve the precision of the estimates, generalised regression (GREG) estimation allows the use of more than one auxiliary variable, and hence has the potential to be more efficient than number-raised and ratio estimation.

2. In order to realise these efficiencies, it will be necessary to develop a methodology that will produce generalised regression estimates and measures of accuracy of these generalised regression estimates for a variety of statistics under various sample designs. The linearization (or Taylor expansion) method and resampling (or replication) methods, such as Jackknife, Balanced Repeated Replications (BRR) and the Bootstrap, are generally used for non-linear statistics, such as ratio or regression estimators. While the linearization method involves the derivation of separate standard error formulae for each non-linear statistic, the resampling methods, although computationally intensive, require a single standard error formulae for all statistics. Hence, it was recommended that the various resampling methods should be evaluated for the ABS business surveys using generalised regression estimation.

3. The choice of an appropriate variance estimation methods for point-in-time and movement estimates will depend to some extent on the underlying sample design methodology. The Grouped Partially Balanced Repeated Replication (GPBRR), the Simple Random Sampling With Replacement (SRSWR) Bootstrap, and the Simple Random Sampling Without Replacement (SRSWOR) Bootstrap options were considered for point-in-time estimates, while the GPBRR, one component SRSWR Bootstrap and SRSWOR Bootstrap, and the three component SRSWR Bootstrap and SRSWOR Bootstrap options were considered for movement estimates.

4. A simulation study was conducted in two stages. The first stage was to compare the various variance estimation methods for number-raised and ratio point-in-time and movement estimates, while the second stage was to evaluate the "best" variance estimation method for GREG point-in-time and movement estimates. The first stage of the simulation study found that the Grouped Partially Balanced Repeated Replication (GPBRR) was not suited to the business survey's situation, and that the "best" variance estimation method is the SRSWOR Bootstrap for point-in-time estimates and the one component SRSWOR Bootstrap for movement estimates. The second stage of the simulation study found that the SRSWOR Bootstrap was also appropriate for GREG point-in-time estimates. At the time of this paper, the findings of the evaluation of the "best" variance estimation method for GREG movement estimates was not available.

5. The recommendation of this paper is that the computer system should use the Bootstrap (where the selection of replicate samples are by SRSWOR) to produce estimated variances of the generalised regression estimates. This method is acceptable when measured against the four characteristics required in the computing system and the underlying methodology mentioned at the start of the paper, with one exception. The exception is that it is not possible to conclude with sufficient power that the generalised regression point-in-time and movement variance estimators are unbiased.

6. The two alternative variance estimation methods are the Jacknife and Balanced Repeated Replication. The Jacknife is unsuitable because it requires many more replicate weights (i.e. the number of replicate weights needs to be at least equal to the sample size in each calibration class) for storage and processing. The Balanced Repeated Replication is unsuitable because of its simulation properties, specifically the efficiency of the variance estimator as measured by the root mean squared error.

7. The questions for MAC members in relation to the developing of a methodology and a computer system that will produce generalised regression estimates and measures of accuracy of these generalised regression estimates are:

- Is there sufficient evidence to support the recommendation to adopt the Bootstrap (SRSWOR)? If not, what should be done to obtain this evidence?
- Is there sufficient theoretical and empirical evidence in the literature to support the hypothesis that the Bootstrap variance estimator is unbiased, or should the simulation study be extended?
- What is the minimum acceptable level of simulation error? (i.e. how many replicate weights should be used?)
- What should be done to determine this acceptable level?
- What brief guidance can be provided on the other issues listed in the conclusion?

Part 2: The GREG Estimator and Variance Estimators

2.1 Introduction

8. The availability of Business Activity Statement (BAS) data collected by the Australian Taxation Office (ATO) has provided the Australian Bureau of Statistics (ABS) with opportunities to improve the efficiency of sample designs and estimations for its business surveys. The ABS business surveys currently use two methods of estimation; number-raised estimation and ratio estimation. While ratio estimation allows the use of one auxiliary variable to improve the precision of the estimates, generalised regression (GREG) estimation allows the use of more than one auxiliary variable, and hence has the potential to be more efficient (i.e. reduce the current sample sizes for ABS business surveys with no reduction in the accuracy of the estimates) than number-raised and ratio estimation.

9. In order to realise these efficiencies, it will be necessary to develop a methodology that will produce generalised regression estimates and measures of accuracy of these generalised regression estimates for a variety of statistics under various sample designs; including:

- single phase point-in-time estimates of level
- single phase movement estimates of level
- single phase point-in-time estimates of rates
- single phase movement estimates of rates
- two phase point-in-time estimates of level
- two phase movement estimates of level
- two phase point-in-time estimates of rates
- two phase movement estimates of rates

10. It will also be necessary to develop a computer system that will produce generalised regression estimates and measures of accuracy of these generalised regression estimates. There are several characteristics that would be required in this computing system, and the underlying methodology. Firstly, the computer system needs to be generic. This means that the algorithm for variance estimation is the same irrespective of the corresponding statistics. This simplifies the maintenance and development of the computing system. Secondly, the computer system needs to minimise the computation costs. These include the storage costs (e.g. the replicate weights in the case of the replicate variance estimators) and the time the computing system takes to calculate the variance estimators. Thirdly, the methodology needs to have desirable theoretical properties, such as unbiasedness. Fourthly, the methodology to have good empirical properties. This includes information about the bias and variance of the method that can be measured in simulation studies. All of these four characteristics can not be considered without consideration of the typical ABS sample design methodology (e.g. studies in the literature that simulation properties of variance estimation methods depend heavily on the sample design methodology).

11. This part of the paper describes the generalised regression estimator for single phase and two-phase sample designs, and the various variance estimation methods for point-in-time and movement estimates for single phase sample designs. At the time of this paper, the various variance estimation methods had not been finalised for point-in-time and movement estimates for two-phase sample designs.

2.2 The Generalised Regression Estimator

12. The generalised regression estimator has several advantages over the standard Horvitz-Thompson estimator. Firstly, the estimates will be consistent with known population totals, and secondly, the estimates will usually be more accurate (for a fixed sample size). Furthermore, the generalised regression estimator is unbiased with respect to the assumed model, and is design consistent.

2.2.1 Generalised Regression Estimator for Single-Phase Sample Designs

13. Consider a finite population $U = \{1,...,i,...,N\}$, from which a probability sample $s\ (s \subseteq U)$ is drawn according to a sample design with selection probabilities $\pi_i = \Pr(i \in s)$. The sampling weights $w_i = 1/\pi_i$ are those used in the Horvitz-Thompson estimator, $\hat{t}_{y\pi} = \sum_{i \in s} w_i y_i$, for variable of interest y. The objective is to estimate the population total $Y = \sum_{i \in U} y_i$, where y_i is the value of the variable of interest y for unit i. Assume there exists a set of auxiliary variables $\underset{\sim}{x}_i = (x_{1i},...,x_{ki},...,x_{Ki})^T$ for which the population totals $\underset{\sim}{t}_x = \sum_{i \in U} \underset{\sim}{x}_i$ are known. The generalised regression estimator is given by:

$$\hat{t}_{yreg} = \sum_{i \in s} w_i y_i + \left(\underset{\sim}{t}_x - \sum_{i \in s} w_i \underset{\sim}{x}_i \right)^T \underset{\sim}{\hat{\beta}}$$

where $\underset{\sim}{\hat{\beta}} = \left(\sum_{i \in s} w_i \underset{\sim}{x}_i \underset{\sim}{x}_i^T \right)^{-1} \left(\sum_{i \in s} w_i \underset{\sim}{x}_i y_i \right) = T^{-1} \left(\sum_{i \in s} w_i \underset{\sim}{x}_i y_i \right)$.

14. The generalised regression estimator is often written as:

$$\hat{t}_{yreg} = \sum_{i \in s} w_i g_i y_i$$

where g_i the g-weight for unit i, defined as:

$$g_i = \left(1 + \underset{\sim}{x}_i^T T^{-1} \left(\underset{\sim}{t}_x - \sum_{i \in s} w_i \underset{\sim}{x}_i \right)^T \right).$$

15. Alternatively, the auxiliary information can be used to produce an unbiased estimator, $\hat{t}_{ycal} = \sum_{i \in s} \tilde{w}_i y_i$, by modifying the weights by a process known as calibration (Deville and Sarndal 1992). A new set of calibrated weights, $\tilde{w}_i$, are sought which lie as close as possible to the set of design weights, w_i. The calibration requires the specification of distance function between the calibration weight and the design weight. Although any one of a number of distance functions could be used, the most commonly used is the generalised least squares distance function:

$$D = \sum_{i \in s} \frac{(\tilde{w}_i - w_i)^2}{w_i}$$

Minimisation of the generalised least squares distance function using Lagrange multipliers, subject to satisfying the calibration constraints, $\sum_{i \in s} \tilde{w}_i \underset{\sim}{x}_i = \underset{\sim}{t}_x$, leads to the calibrated weights:

$$\tilde{w}_i = w_i \left(1 + \underset{\sim}{x}_i^T T^{-1} \left(\underset{\sim}{t}_x - \underset{\sim}{\hat{t}}_{x\pi} \right)^T \right)$$

where $\underset{\sim}{\hat{t}}_{x\pi} = \sum_{i \in s} w_i \underset{\sim}{x}_i$ is the Horvitz-Thompson estimator for the set of auxiliary information.

16. The calibrated estimator is equivalent to the generalised regression estimator:

$$\hat{t}_{ycal} = \sum_{i \in s} \tilde{w}_i y_i = \sum_{i \in s} w_i \left(1 + \underset{\sim}{x}_i^T T^{-1} \left(\underset{\sim}{t}_x - \underset{\sim}{\hat{t}}_{x\pi} \right)^T \right) y_i = \sum_{i \in s} w_i g_i y_i = \hat{t}_{yreg}$$

However, one disadvantage of the generalised regression estimator is that it can give negative weights, while under the calibrated estimator it is possible to impose range restrictions on the calibrated weights, $L \le g_i \le U$, where L and U are suitable lower and upper bounds. In order to satisfy the benchmark constraints and the range restrictions, the calculation of the calibration weights need to be undertaken using an iterative method (Singh and Mohl 1996).

2.2.1 Generalised Regression Estimator for Two-Phase Sample Designs

17. Consider a finite population $U = \{1, \ldots, i, \ldots, N\}$, from which a first-phase probability sample $s_1 \; (s_1 \subseteq U)$ is drawn according to a sample design with selection probabilities $\pi_{1i} = \Pr(i \in s_1)$. A second-phase probability sample $s_2 \; (s_2 \subseteq s_1 \subseteq U)$ is drawn from the first-phase sample according to a sample design with selection probabilities $\pi_{2i} = \Pr(i \in s_2 | s_1)$. The first-phase and second-phase sampling weights $w_{1i} = 1/\pi_{1i}$ and $w_{2i} = 1/\pi_{2i}$ are those used in the two-phase estimator, $\hat{t}_{y\pi} = \sum_{i \in s_2} w_{1i} w_{2i} y_i = \sum_{i \in s_2} w_i y_i$, for second-phase variable of interest y, where $w_i = w_{1i} w_{2i}$.

18. The objective is to estimate the population total $Y = \sum_{i \in U} y_i$, where y_i is the value of the variable of interest y for unit i. Assume there exists two sets of auxiliary variables $\underset{\sim}{x}_{1i}$ and $\underset{\sim}{x}_{2i}$. The population totals $\underset{\sim}{t}_{x_1} = \sum_{i \in U} \underset{\sim}{x}_{1i}$ are known for $\underset{\sim}{x}_{1i}$, whereas the population totals are unknown for $\underset{\sim}{x}_{2i}$, but the auxiliary variables are available for the first-phase sample.

19. The two sets of auxiliary information can be used to produce an approximately unbiased estimator, $\hat{t}_{ycal} = \sum_{i \in s_2} \tilde{w}_i y_i$, by modifying the weights using calibration. The first step is to minimise the generalised least squares distance function:

$$D_1 = \sum_{i \in s_1} \frac{\left(\tilde{w}_{1i} - w_{1i}\right)^2}{w_{1i}}$$

subject to satisfying the first-phase calibration constraints, $\sum_{i \in s_1} \tilde{w}_i \underset{\sim}{x}_{1i} = \underset{\sim}{t}_{x_1}$. The minimisation of the generalised least squares distance function, using Lagrange multipliers, leads to the first-phase calibrated weights:

$$\tilde{w}_{1i} = w_{1i}\left(1 + \underset{\sim}{x}_{1i}^{T} T_1^{-1}\left(\underset{\sim}{t}_{x_1} - \hat{\underset{\sim}{t}}_{x_1\pi}\right)^{T}\right) = w_{1i} g_{1i}$$

where $\hat{\underset{\sim}{t}}_{x_1\pi_1} = \sum_{i \in s_1} w_{1i} \underset{\sim}{x}_{1i}$ is the first-phase estimator for the first set of auxiliary information.

20. The second step is to minimise the generalised least squares distance function:

$$D_2 = \sum_{i \in s_2} \frac{\left(\tilde{w}_i - \tilde{w}_{1i} w_{2i}\right)^2}{\tilde{w}_{1i} w_{2i}}$$

subject to satisfying the second-phase calibration constraints, $\sum_{i \in s_2} \tilde{w}_i \underset{\sim}{x}_i = \sum_{i \in s_1} \tilde{w}_{1i} \underset{\sim}{x}_i$, where $\underset{\sim}{x}_i = \left(\underset{\sim}{x}_{1i}, \underset{\sim}{x}_{2i}\right)$. The minimisation of the generalised least squares distance function, using Lagrange multipliers, leads to the second-phase calibrated weights:

$$\tilde{w}_i = \tilde{w}_{1i} w_{2i}\left(1 + \underset{\sim}{x}_i^{T} T_2^{-1}\left(\hat{\underset{\sim}{t}}_{x\pi_1} - \hat{\underset{\sim}{t}}_{x\pi_2}\right)^{T}\right) = w_{1i} g_{1i} w_{2i} g_{2i} = w_i g_i$$

where $\hat{\underset{\sim}{t}}_{x\pi_1} = \sum_{i \in s_1} \tilde{w}_{1i} \underset{\sim}{x}_i$ is the first-phase estimator for the set of auxiliary information, and $\hat{\underset{\sim}{t}}_{x\pi_2} = \sum_{i \in s_2} \tilde{w}_{1i} w_{2i} \underset{\sim}{x}_i$ is second-phase estimator for the set of auxiliary information.

21. An alternative approach (Hidiroglou, Estevao and Arcaro 2000) is to use the second-phase design weights $w_i = w_{1i}w_{2i}$ as the starting weights in the second-phase calibration, rather than using $\tilde{w}_{1i}w_{2i}$ as the starting weights, and hence minimise the generalised least squares distance function:

$$D_2 = \sum_{i \in s_2} \frac{\left(\tilde{w}_i - w_{1i}w_{2i}\right)^2}{w_{1i}w_{2i}}$$

subject to satisfying the second-phase calibration constraints, $\sum_{i \in s_2} \tilde{w}_i \underset{\sim}{x}_i = \sum_{i \in s_1} \tilde{w}_{1i} \underset{\sim}{x}_i$, where $\underset{\sim}{x}_i = \left(\underset{\sim}{x}_{1i}, \underset{\sim}{x}_{2i}\right)$. The minimisation of the generalised least squares distance function, using Lagrange multipliers, leads to the second-phase calibrated weights:

$$\tilde{w}_i = w_{1i}w_{2i}\left(1 + \underset{\sim}{x}_i^T T_2^{-1}\left(\hat{\underset{\sim}{t}}_{x\pi_1} - \hat{\underset{\sim}{t}}_{x\pi_2}\right)^T\right) = w_{1i}w_{2i}g_{2i} = w_i g_{2i}$$

22. During the first-phase and second-phase calibration it is possible to impose range restrictions on the first-phase and second-phase calibrated weights, $L_1 \le g_{1i} \le U_1$ and $L_2 \le g_{2i} \le U_2$, where L and U are suitable lower and upper bounds.

23. While these two alternative distance measures produce slightly different estimators, they have very similar properties. Since there was no obvious choice based on methodological reasons, the second method was chosen based on computer system efficiency reasons.

2.3 Variance Estimators

24. The linearization (or Taylor expansion) method is often used for non-linear statistics, such as ratio or regression estimators, although resampling (or replication) methods, such as Jackknife, Balanced Repeated Replications (BRR) and the Bootstrap, are becoming more fashionable. While the linearization method is applicable to general sampling designs, the major disadvantage of this method is that it involves the derivation of separate standard error formulae for each non-linear statistic and hence additional computer programming. On the other hand, the resampling methods require a single standard error formulae for all statistics, but the major disadvantage of these methods is that they are computationally expensive. With the availability of more powerful computers, it is thought that the disadvantages associated with the resampling methods are minor compared to those associated with linearization methods. Hence, it was recommended that the various resampling methods should be evaluated for the ABS business surveys using generalised regression estimation.

25. While the Jackknife is considered to be the most accurate of these resampling methods, it is also considered to be more costly and less timely in terms of processing than the other resampling methods (Shao and Tu 1995). Furthermore, the Jackknife requires many more replicate weights for storage and processing. Under the current estimation system for ABS business surveys, efficient computing procedures have been used to implement the Jackknife in order to substantially reduce these costs and improve the timeliness. Unfortunately, these procedures are only applicable for number-raised and ratio estimates and cannot be extended for use with generalised regression estimates. Hence, it was decided that the Jackknife would be excluded from the evaluation.

26. The derivation of the point-in-time and movement variance estimators involved two steps. Firstly, the variance for the number raised estimator are explicitly derived. Secondly, an unbiased estimator of the number raised estimator with respect to the replicate sampling methodology are derived (e.g. SRSWR or SRSWOR). The derivation of the point-in-time and movement variance estimators extends to generalised regression estimation using a Taylor Series argument: the expectation of the replicate variance estimator, with respect to resampling, is equal to the Taylor Series variance estimator. (Note: the Taylor Series variance estimator is a linear combination of number raised variance estimates.)

2.3.1 Variance Estimation Methods for Point-in-time Estimators

27. The choice of an appropriate variance estimation method for point-in-time estimates will depend to some extent on the underlying sample design methodology. The majority of ABS business surveys are designed to produce reliable estimates for a number of variables across a number of classifications. The accuracy of the estimates for the classifications are generally controlled by stratifying the survey frame by these classifications (in the case where the classifications are available for all sampling units on the survey frame). The accuracy of the estimates are further improved by stratifying the sampling frame by other auxiliary variables related to the variables of interest. Therefore, the majority of ABS business surveys consist of a very large number of strata.

Balanced Repeated Replications

28. The standard procedure for the balanced repeated replications (BRR) is not well suited to the situation where there are a large number of strata, due to the cost of processing the large number of balanced replicates. In this situation it is possible to use a set of partially balanced replicates by dividing the large number of strata into a smaller number of groups (Wolter 1985). The procedure for the grouped partially balanced repeated replications (GPBRR) for a stratified random sample selected without replacement is to:

(i) Randomly divide the original sample of n_h units into two equal groups of $m_h = n_h/2$ units, independently within each stratum h.

(ii) Divide the total number of strata, H, into G groups with an equal number of strata within each group.

(iii) Construct a set of R balanced half samples (replications) using Hadamard matrices, where R is a multiple of four greater than or equal to G.

(iv) The GPBRR sampling weights within each balanced half sample r are given by:

$$w_i^{(r)} = \begin{cases} w_i\left(1+\sqrt{\dfrac{(1-f_h)(n_h-m_h)}{m_h}}\right) & ,\text{if } i \in r \\ w_i\left(1-\sqrt{\dfrac{(1-f_h)m_h}{(n_h-m_h)}}\right) & ,\text{if } i \notin r \end{cases}$$

where $f_h = n_h/N_h$ is the sampling fraction within stratum h.

(v) Calculate the GPBRR estimators, $\hat{\theta}_y^{(1)},\ldots,\hat{\theta}_y^{(R)}$, using these sampling weights:

$$\hat{\theta}_y^{(r)} = \sum_{i \in s} w_i^{(r)} y_i$$

(vi) The GPBRR variance estimator is given by:

$$Var\left(\hat{\theta}_y\right) = \frac{\sum_{r=1}^{R}\left(\hat{\theta}_y^{(r)} - \bar{\hat{\theta}}_y\right)^2}{R}$$

where $\bar{\hat{\theta}}_y = \sum_{r=1}^{R} \hat{\theta}_y^{(r)} \Big/ R$.

Bootstrap

29. The standard bootstrap procedure for a stratified random sample selected without replacement (Shao and Tu 1995) is to:

(i) Select a simple random sample of m_h units with replacement from the original sample of n_h units, independently within each stratum h. Let r_{hi}^* denote the number of times unit i in stratum h is included. The bootstrap sampling weights are given by:

$$w_i^* = w_i \left(1 - \sqrt{\frac{(1-f_h)m_h}{(n_h-1)}} + \sqrt{\frac{(1-f_h)m_h}{(n_h-1)}} \frac{n_h}{m_h} r_{hi}^* \right)$$

where $f_h = n_h / N_h$ is the sampling fraction within stratum h.

Alternatively, if the simple random sample of m_h units is selected without replacement from the original sample of n_h units, the bootstrap sampling weights are given by:

$$w_i^* = w_i \left(1 - \sqrt{\frac{(1-f_h)m_h}{(n_h-m_h)}} + \sqrt{\frac{(1-f_h)m_h}{(n_h-m_h)}} \frac{n_h}{m_h} r_{hi}^* \right)$$

The bootstrap estimator is calculated using these bootstrap sampling weights:

$$\hat{\theta}_y^* = \sum_{i \in s} w_i^* y_i$$

(ii) Independently replicate step (i) a large number of times, B, and calculate the bootstrap estimates $\hat{\theta}_y^{(1)}, \ldots, \hat{\theta}_y^{(B)}$.

(iii) The Monte Carlo approximation to the bootstrap variance estimator is given by:

$$Var\left(\hat{\theta}_y\right) = \frac{\sum_{b=1}^{B} \left(\hat{\theta}_y^{(b)} - \bar{\hat{\theta}}_y \right)^2}{B-1}$$

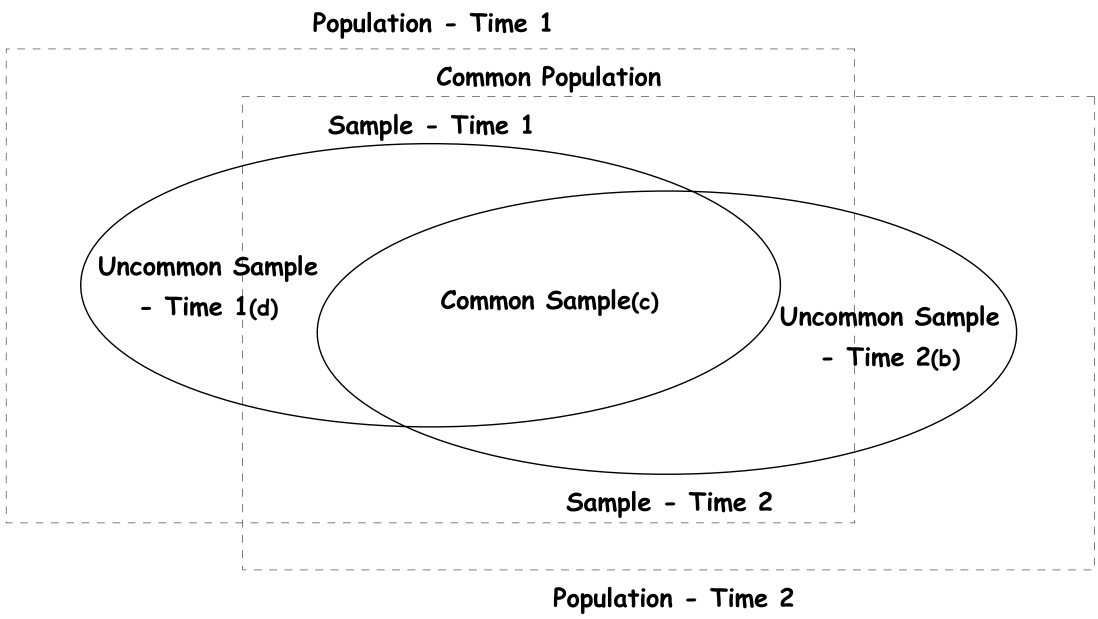
where $\bar{\hat{\theta}}_y = \sum_{b=1}^{B} \hat{\theta}_y^{(b)} \Big/ B$.

2.3.2 Variance Estimation Methods for Movement Estimators

30. The choice of an appropriate variance estimation method for movement estimates will also depend to some extent on the sample design methodology used to produce the movement estimates. Most ABS repeated business surveys are designed to produce reliable point-in-time and movement estimates. The reliability of the point-in-time estimates are generally controlled through the allocation of the samples, while the reliability of the movement estimates are generally controlled through both the allocation and selection of the samples. Another objective of ABS repeated business surveys is to reduce provider load on individual businesses, and hence these surveys have also been designed to minimise the number of small and medium businesses (through sample allocation) as well as control the rotation of selected small and medium business (through sample selection). The sample selections for most ABS repeated business surveys are generally designed to enable a balance between reducing the compliance cost (i.e. provider load) on individual businesses and producing reliable movement estimates. The sample selections for many ABS repeated business surveys are performed using synchronised sampling (Brewer, Gross and Lee 2000).

31. The derivation of the various movement variance estimators under the sampling methodology used for ABS repeated business survey is best understood by means of a pictorial illustration (Figure 1).

Figure 1: Pictorial Illustration of Sampling Methodology for ABS Repeated Business Surveys



32. The units selected in the survey can be partitioned into three categories: category (d) units are selected in the survey at time point 1 but not selected in the survey at time point 2; category (c) units are selected in the survey at time points 1 and 2; and category (b) units are selected in the survey at time point 2 but not selected in the survey at time point 1. The derivation of variance estimation methods

for movement estimates under this sample design methodology have not been considered in the literature.

Balanced Repeated Replications

33. The procedure for the grouped partially balanced repeated replications (GPBRR) for a stratified random sample selected without replacement is to:

(i) Divide the units selected in the survey into the three categories (d,c,b) described above.

(ii) Within each of the three categories (d,c,b) randomly divide the original samples of n_{1uh}, n_{ch}, n_{2uh} units into two equal groups of $m_{1uh} = n_{1uh}/2, m_{ch} = n_{ch}/2, m_{2uh} = n_{2uh}/2$ units, independently within each stratum h, where n_{1uh}, n_{ch}, n_{2uh} are the numbers of units selected in the survey in the three categories (d,c,b).

(iii) Divide the total number of strata, H, into G groups with an equal number of strata within each group.

(iv) Construct a set of R balanced half samples (replications) using Hadamard matrices, where R is a multiple of four greater than or equal to G.

(v) The GPBRR sampling weights within each balanced half sample r are given by:

$$w_{1i}^{(r)} = \begin{cases} w_{1i}\left(1+\sqrt{\dfrac{(1-f_{ch})(n_{ch}-m_{ch})}{m_{ch}}}\right) & \text{,if } i \in (c) \text{ and } i \in r \\ w_{1i}\left(1-\sqrt{\dfrac{(1-f_{ch})m_{ch}}{(n_{ch}-m_{ch})}}\right) & \text{,if } i \in (c) \text{ and } i \notin r \\ w_{1i}\left(1+\sqrt{\dfrac{\left[n_{1h}(1-f_{1h})-n_{ch}(1-f_{ch})\right](n_{1uh}-m_{1uh})}{m_{1uh}n_{1uh}}}\right) & \text{,if } i \in (d) \text{ and } i \in r \\ w_{1i}\left(1-\sqrt{\dfrac{\left[n_{1h}(1-f_{1h})-n_{ch}(1-f_{ch})\right]m_{1uh}}{(n_{1uh}-m_{1uh})n_{1uh}}}\right) & \text{,if } i \in (d) \text{ and } i \notin r \end{cases}$$

$$w_{2i}^{(r)} = \begin{cases} w_{2i}\left(1+\sqrt{\dfrac{(1-f_{ch})(n_{ch}-m_{ch})}{m_{ch}}}\right) & \text{,if i} \in (c) \text{ and i} \in r \\ w_{2i}\left(1-\sqrt{\dfrac{(1-f_{ch})m_{ch}}{(n_{ch}-m_{ch})}}\right) & \text{,if i} \in (c) \text{ and i} \notin r \\ w_{2i}\left(1+\sqrt{\dfrac{\left[n_{2h}(1-f_{2h})-n_{ch}(1-f_{ch})\right](n_{2uh}-m_{2uh})}{m_{2uh}n_{2uh}}}\right) & \text{,if i} \in (b) \text{ and i} \in r \\ w_{2i}\left(1-\sqrt{\dfrac{\left[n_{2h}(1-f_{2h})-n_{ch}(1-f_{ch})\right]m_{2uh}}{(n_{2uh}-m_{2uh})n_{2uh}}}\right) & \text{,if i} \in (b) \text{ and i} \notin r \end{cases}$$

where $f_{ch} = n_{1ch}n_{2ch}/n_{ch}N_{ch}$.

(vi) Calculate the GPBRR estimators, $\hat{\theta}_{ym}^{(1)},\dots,\hat{\theta}_{ym}^{(R)}$, using these sampling weights:

$$\hat{\theta}_{ym}^{(r)} = \sum_{i\in(d,c)} w_{2i}^{(r)}y_{2i} - \sum_{i\in(b,c)} w_{1i}^{(r)}y_{1i}$$

(vii) The GPBRR variance estimator is given by:

$$Var\left(\hat{\theta}_{ym}\right) = \frac{\sum_{r=1}^{R}\left(\hat{\theta}_{ym}^{(r)} - \bar{\hat{\theta}}_{ym}\right)^2}{R}$$

where $\bar{\hat{\theta}}_{ym} = \sum_{r=1}^{R}\hat{\theta}_{ym}^{(r)} \Big/ R$.

Bootstrap

34. The proposed bootstrap procedure for stratified random samples selected without replacement is to:

(i) Divide the units selected in the survey can be partitioned into the three categories (d,c,b) described above.

(ii) Within each of the three categories (d,c,b) select simple random samples of m_{1uh}, m_{ch}, m_{2uh} units with replacement from the original samples of n_{1uh}, n_{ch}, n_{2uh} units, independently within each stratum h, where n_{1uh}, n_{ch}, n_{2uh} are the numbers of units selected in the survey in the three categories (d,c,b).

Let r_{1hi}^{*} and r_{2hi}^{*} denote the number of times unit i in stratum h is included at time points 1 and 2. The bootstrap sampling weights are given by:

$$w_{1i}^{*}=\begin{cases} w_{1i}\left(1-\sqrt{\dfrac{(1-f_{ch})m_{ch}}{(n_{ch}-1)}}+\sqrt{\dfrac{(1-f_{ch})m_{ch}}{(n_{ch}-1)}}\dfrac{n_{ch}}{m_{ch}}r_{1hi}^{*}\right) & ,\text{if } i\in(c) \\ w_{1i}\left(\begin{array}{l} 1-\sqrt{\dfrac{\left[n_{1h}(1-f_{1h})-n_{ch}(1-f_{ch})\right]m_{1uh}}{n_{1uh}(n_{1uh}-1)}} \\ +\sqrt{\dfrac{\left[n_{1h}(1-f_{1h})-n_{ch}(1-f_{ch})\right]m_{1uh}}{n_{1uh}(n_{1uh}-1)}}\dfrac{n_{1uh}}{m_{1uh}}r_{1hi}^{*} \end{array}\right) & ,\text{if } i\in(d) \end{cases}$$

$$w_{2i}^{*}=\begin{cases} w_{2i}\left(1-\sqrt{\dfrac{(1-f_{ch})m_{ch}}{(n_{ch}-1)}}+\sqrt{\dfrac{(1-f_{ch})m_{ch}}{(n_{ch}-1)}}\dfrac{n_{ch}}{m_{ch}}r_{2hi}^{*}\right) & ,\text{if } i\in(c) \\ w_{2i}\left(\begin{array}{l} 1-\sqrt{\dfrac{\left[n_{2h}(1-f_{2h})-n_{ch}(1-f_{ch})\right]m_{2uh}}{n_{2uh}(n_{2uh}-1)}} \\ +\sqrt{\dfrac{\left[n_{2h}(1-f_{2h})-n_{ch}(1-f_{ch})\right]m_{2uh}}{n_{2uh}(n_{2uh}-1)}}\dfrac{n_{2uh}}{m_{2uh}}r_{2hi}^{*} \end{array}\right) & ,\text{if } i\in(b) \end{cases}$$

where $f_{ch}=n_{1ch}n_{2ch}/n_{ch}N_{ch}$.

Alternatively, if the simple random samples of m_{1uh}, m_{ch}, m_{2uh} units are selected without replacement from the original samples of n_{1uh}, n_{ch}, n_{2uh} units, the bootstrap sampling weights are given by:

$$w_{1i}^{*}=\begin{cases} w_{1i}\left(1-\sqrt{\dfrac{(1-f_{ch})m_{ch}}{(n_{ch}-m_{ch})}}+\sqrt{\dfrac{(1-f_{ch})m_{ch}}{(n_{ch}-m_{ch})}}\dfrac{n_{ch}}{m_{ch}}r_{1hi}^{*}\right) & ,\text{if } i\in(c) \\ w_{1i}\left(\begin{array}{l} 1-\sqrt{\dfrac{\left[n_{1h}(1-f_{1h})-n_{ch}(1-f_{ch})\right]m_{1uh}}{n_{1uh}(n_{1uh}-m_{1uh})}} \\ +\sqrt{\dfrac{\left[n_{1h}(1-f_{1h})-n_{ch}(1-f_{ch})\right]m_{1uh}}{n_{1uh}(n_{1uh}-m_{1uh})}}\dfrac{n_{1uh}}{m_{1uh}}r_{1hi}^{*} \end{array}\right) & ,\text{if } i\in(d) \end{cases}$$

$$w_{2i}^{*} = \begin{cases} w_{2i}\left(1-\sqrt{\dfrac{(1-f_{ch})m_{ch}}{(n_{ch}-m_{ch})}}+\sqrt{\dfrac{(1-f_{ch})m_{ch}}{(n_{ch}-m_{ch})}}\dfrac{n_{ch}}{m_{ch}}r_{2hi}^{*}\right) & \text{,if } i\in(c) \\ w_{2i}\left(\begin{array}{l} 1-\sqrt{\dfrac{\left[n_{2h}(1-f_{2h})-n_{ch}(1-f_{ch})\right]m_{2uh}}{n_{2uh}(n_{2uh}-m_{2uh})}} \\ +\sqrt{\dfrac{\left[n_{2h}(1-f_{2h})-n_{ch}(1-f_{ch})\right]m_{2uh}}{n_{2uh}(n_{2uh}-m_{2uh})}}\dfrac{n_{2uh}}{m_{2uh}}r_{2hi}^{*} \end{array}\right) & \text{,if } i\in(b) \end{cases}$$

The bootstrap estimator is calculated using these bootstrap sampling weights:

$$\hat{\theta}_{ym}^{*} = \sum_{i\in(d,c)} w_{2i}^{*}y_{2i} - \sum_{i\in(b,c)} w_{1i}^{*}y_{1i}$$

(iii) Independently replicate step (ii) a large number of times, B, and calculate the bootstrap estimates $\hat{\theta}_{ym}^{(1)},\dots,\hat{\theta}_{ym}^{(B)}$.

(iv) The Monte Carlo approximation to the bootstrap variance estimator is given by:

$$Var\left(\hat{\theta}_{ym}\right) = \frac{\sum_{b=1}^{B}\left(\hat{\theta}_{ym}^{(b)}-\bar{\hat{\theta}}_{ym}\right)^{2}}{B-1}$$

where $\bar{\hat{\theta}}_{ym} = \sum_{b=1}^{B}\hat{\theta}_{ym}^{(b)} \Big/ B$.

35. An alternative to the bootstrap procedure described above is to split the variance estimator into three components: the variance at time 1; the variance at time 2; and the covariance between time 1 and time 2.

$$Var\left(\hat{\theta}_{ym}\right) = Var\left(\hat{\theta}_{y1}\right) + Var\left(\hat{\theta}_{y2}\right) - 2\times Cov\left(\hat{\theta}_{y1},\hat{\theta}_{y2}\right)$$

36. The alternative bootstrap procedure for stratified random samples selected without replacement is to:

(i) Divide the units selected in the survey into the three categories (d,c,b) described above.

(ii) Within each of the three categories (d,c,b) select simple random samples of m_{1uh}, m_{ch}, m_{2uh} units with replacement from the original samples of n_{1uh}, n_{ch}, n_{2uh} units, independently within each stratum h, where n_{1uh}, n_{ch}, n_{2uh} are the numbers of units selected in the survey in the three categories (d,c,b), where:

$$m_{1uh} = n_{1uh} \times \min\left\{(n_{1h}-2)/n_{1h}, (n_{2h}-2)/n_{2h}\right\}$$

$$m_{ch} = n_{ch} \times \min\left\{(n_{1h}-2)/n_{1h}, (n_{2h}-2)/n_{2h}\right\}$$

$$m_{2uh} = n_{2uh} \times \min\left\{(n_{1h}-2)/n_{1h}, (n_{2h}-2)/n_{2h}\right\}$$

Let r^*_{1hi} and r^*_{2hi} denote the number of times unit i in stratum h is included at time points 1 and 2. The bootstrap sampling weights are given by:

$$w^*_{1i} = w_{1i}\left(1 - \sqrt{\frac{(1-f_{1h})m_{1h}}{(n_{1h}-1)}} + \sqrt{\frac{(1-f_{1h})m_{1h}}{(n_{1h}-1)}}\frac{n_{1h}}{m_{1h}} r^*_{1hi}\right) \quad \text{,if } i \in (d,c)$$

$$w^*_{2i} = w_{2i}\left(1 - \sqrt{\frac{(1-f_{2h})m_{2h}}{(n_{2h}-1)}} + \sqrt{\frac{(1-f_{2h})m_{2h}}{(n_{2h}-1)}}\frac{n_{2h}}{m_{2h}} r^*_{2hi}\right) \quad \text{,if } i \in (b,c)$$

$$w^*_{1ci} = w_{1i}\left(1 - \sqrt{\frac{(1-f_{ch})m_{ch}}{(n_{ch}-1)}} + \sqrt{\frac{(1-f_{ch})m_{ch}}{(n_{ch}-1)}}\frac{n_{ch}}{m_{ch}} r^*_{1hi}\right) \quad \text{,if } i \in (c)$$

$$w^*_{2ci} = w_{2i}\left(1 - \sqrt{\frac{(1-f_{ch})m_{ch}}{(n_{ch}-1)}} + \sqrt{\frac{(1-f_{ch})m_{ch}}{(n_{ch}-1)}}\frac{n_{ch}}{m_{ch}} r^*_{2hi}\right) \quad \text{,if } i \in (c)$$

where $m_{1h} = m_{1uh} + m_{ch}$ and $m_{2h} = m_{2uh} + m_{ch}$.

The bootstrap estimators are calculated using these bootstrap sampling weights:

$$\hat{\theta}^*_{y1} = \sum_{i \in (d,c)} w^*_{1i} y_{1i}$$

$$\hat{\theta}^*_{y2} = \sum_{i \in (b,c)} w^*_{2i} y_{2i}$$

$$\hat{\theta}^{*}_{y1c} = \sum_{i \in (c)} w^{*}_{1ci} y_{1i}$$

$$\hat{\theta}^{*}_{y2c} = \sum_{i \in (c)} w^{*}_{1ci} y_{1i}$$

(iii) Independently replicate step (ii) a large number of times, B, and calculate the bootstrap estimates $\hat{\theta}^{(1)}_{y1},\ldots,\hat{\theta}^{(B)}_{y1}$, $\hat{\theta}^{(1)}_{y2},\ldots,\hat{\theta}^{(B)}_{y2}$, $\hat{\theta}^{(1)}_{y1c},\ldots,\hat{\theta}^{(B)}_{y1c}$ and $\hat{\theta}^{(1)}_{y2c},\ldots,\hat{\theta}^{(B)}_{y2c}$.

(iv) The Monte Carlo approximation to the bootstrap variance estimator is given by:

$$Var\left(\hat{\theta}_{ym}\right) = \frac{\sum_{b=1}^{B}\left(\hat{\theta}^{(b)}_{y1} - \overline{\hat{\theta}}_{y1}\right)^2}{B-1} + \frac{\sum_{b=1}^{B}\left(\hat{\theta}^{(b)}_{y2} - \overline{\hat{\theta}}_{y2}\right)^2}{B-1} - 2 \times \frac{\sum_{b=1}^{B}\left(\hat{\theta}^{(b)}_{y1c} - \overline{\hat{\theta}}_{y1c}\right)\left(\hat{\theta}^{(b)}_{y2c} - \overline{\hat{\theta}}_{y2c}\right)}{B-1}$$

3.1 The Simulation Study

37. The simulation study was conducted in two stages. The first stage was to compare the various variance estimation methods for number-raised and ratio point-in-time and movement estimates, while the second stage was to evaluate the "best" variance estimation method for GREG point-in-time and movement estimates. The first stage was conducted using number-raised and ratio estimates, rather than GREG estimates, since it was possible to compare the variances under the various estimation methods against the standard linearization variance using number-raised and ratio estimates. It was also assumed that the "best" variance estimation method for number-raised and ratio point-in-time and movement estimates would also be the "best" variance estimation method for GREG point-in-time and movement estimates.

38. The simulation study was restricted to single phase point-in-time estimates of level and single phase movement estimates of level. The more complex point-in-time estimation statistics (e.g. single phase estimates of rates) will not be included in this evaluation, since it is considered that the "best" variance estimation method for single phase point-in-time level estimates will most likely be the "best" variance estimation method for the point-in-time estimates for the more complex point-in-time estimation statistics.

39. The various variance estimation methods were assessed by comparing the relative root mean squared error percents. The relative root mean squared error percents were calculated as the square root of the estimated mean squared error (i.e. estimated bias squared plus estimated variance) of the various variance estimation methods divided by the standard linearization variance (expressed as a percentage). The estimated biases and estimated variances, and hence these estimated mean squared errors, were calculated using one hundred simulations of the variances under the various variance estimation methods:

$$\widehat{RMSE}\left(Var\left(\hat{\theta}_y | s\right)\right) = \frac{\sqrt{\dfrac{\sum_{s=1}^{100}\left(\widehat{Var}\left(\hat{\theta}_y^{(s)} | s\right) - Var\left(\hat{\theta}_y | s\right)\right)^2}{100}}}{Var\left(\hat{\theta}_y | s\right)} \times 100\%$$

where $Var\left(\hat{\theta}_y | s\right)$ is the standard linearization variance given the realised sample s and $Var\left(\hat{\theta}_y^{(i)} | s\right)$ is the estimated variance under the various variance estimation methods for simulation i given the realised sample s.

40. At face value the comparison of the relative root mean squared error percents for the various variance estimation methods might appear to be quite unusual, since it is expected that:

$$\lim_{R\to\infty}\left\{\widehat{RMSE}\left(Var\left(\hat{\theta}_y | s\right)\right)\right\} \to 0$$

for all the various variance estimation methods, where R is the number of replicate samples. Therefore, it is expected that all the various variance estimation methods will be equivalent, under this relative root mean squared error criteria, for very large numbers of replicate samples. However, in practice the number of replicate samples will need to be kept to a minimum in order to keep the computer processing time and costs, as well as computer storage size and costs, to manageable levels.

41. It should also be noted that the results of the simulation study compare the "resampling error" of the various variance estimation methods, and hence do not take into account "sampling error", since the results are based on a single original sample from selected from the population.

3.2 The Results of the Simulation Study

42. The first stage was to compare the various variance estimation methods for point-in-time and movement of number-raised and ratio estimates. The relative root mean squared error percents, under the various variance estimation methods, based on approximately 50 and 100 replicate samples, for the point-in-time number-raised estimates of sales from the "Quarterly Economic Activity Survey, March 2000" are presented in Table 1.

43. While the estimated relative root mean squared error percents for the Grouped Partially Balanced Repeated Replications (GPBRR) were only slightly higher than those for the two Bootstrap options at the Australian level, they were considerably higher at the industry level. Furthermore, as the number of replicates increased from 48 to 96, there appeared to be little improvement in the estimated relative root mean squared error percents at the industry level for the GPBRR. These industry level results were not unexpected, since when the number of strata in an industry is less than 48 the GPBRR estimators with 48 and 96 replicates are equivalent. (The relative root mean squared error percents in Table 1 were not always exactly the same because the replicates groups were formed using different random starts). The GPBRR estimators with 48 and 96 replicates were not the same when the number of strata in an industry is greater than 48 (i.e. Industry 3 and 7). In these industries, it might be expected that the GPBRR estimator with 96 replicates will have much lower simulation error (as measured by relative root mean squared error percents) but this was not the case. It was decided that the GPBRR was probably not suited to the business survey's situation, and hence the GPBRR method was excluded from the remainder of the simulation study.

Table 1: Relative Root Mean Squared Error Percents for Point-in-time Number-Raised Estimates of Sales, under Various Variance Estimation Methods, Quarterly Economic Activity Survey, March 2000

Industry	GPBRR		Bootstrap SRSWR#		Bootstrap SRSWOR#	
	48	96	50	100	50	100
2	88.99	88.99	21.03	14.81	18.25	12.43
3	15.73	15.58	21.41	15.39	10.31	6.92
4	124.92	124.92	21.81	14.06	16.32	11.15
5	34.48	34.33	19.17	14.60	15.77	12.08
6	45.38	45.40	19.97	15.54	18.08	12.92
7	33.83	30.94	19.64	14.59	16.66	13.01
8	38.63	35.76	19.93	15.75	18.58	13.26
9	32.42	32.42	20.78	13.99	20.07	13.45
10	54.32	54.32	22.11	17.26	17.07	12.05
11	55.09	55.16	20.23	13.97	17.81	13.40
12	56.21	56.06	21.06	16.96	17.22	12.31
14	0.00	0.00	0.00	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00
16	26.47	26.41	19.54	13.95	21.84	15.55

17	51.98	51.98	20.78	15.17	19.52	14.89
Total	**29.76**	**29.21**	**20.10**	**15.40**	**18.74**	**12.46**

The SRSWR Bootstrap was based on sample sizes of (m = n - 1), while the SRSWOR Bootstrap was based on sample sizes of (m = n/2).

44. The estimated relative root mean squared error percents for the SRSWOR Bootstrap were consistently slightly lower than those for the SRSWR Bootstrap. An evaluation into an SRSWR Bootstrap based on larger sample sizes of (m = 5n) found that there was little improvement over the SRSWR Bootstrap based on sample sizes of (m = n - 1), and it was still not as good as the SRSWOR Bootstrap based on sample sizes of (m = n/2).

45. The relative root mean squared error percents, under the various variance estimation methods, based on approximately 50 and 100 replicate samples, for the movement of number-raised estimates of sales from the "Quarterly Economic Activity Survey, March and June 2000" are presented in Table 2.

46. The estimated relative root mean squared error percents for the three component SRSWR Bootstrap were consistently higher than the one component SRSWR Bootstrap and SRSWOR Bootstrap options. There was not much difference between the estimated relative root mean squared error percents for the SRSWR Bootstrap and SRSWOR Bootstrap options.

Table 2: Relative Root Mean Squared Error Percents for Movement of Number-Raised Estimates of Sales, under Various Variance Estimation Methods, Quarterly Economic Activity Survey, March and June 2000

Industry	Bootstrap SRSWR(3)#		Bootstrap SRSWR(1)#		Bootstrap SRSWOR(1)#	
	50	**100**	**50**	**100**	**50**	**100**
2	19.45	13.34	20.23	19.20	19.42	16.45
3	22.94	15.24	26.74	14.76	11.58	6.93
4	19.33	11.83	39.03	38.27	40.29	38.86
5	17.43	12.72	22.23	19.12	17.68	12.14
6	27.41	18.53	23.66	14.23	15.21	10.71
7	29.09	24.68	19.57	15.21	21.79	15.83
8	31.59	29.72	29.48	24.41	21.66	19.75
9	24.05	19.82	22.57	17.21	20.89	13.90
10	50.62	34.62	49.37	43.62	49.42	44.37
11	20.82	14.21	19.32	15.47	23.76	16.52
12	22.17	15.17	21.50	15.51	20.37	14.57
14	0.00	0.00	0.00	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00
16	21.71	13.17	23.71	15.71	19.10	11.95
17	33.05	28.26	24.21	16.27	18.79	12.13
Total	**22.79**	**17.86**	**21.67**	**13.82**	**18.27**	**12.86**

[#] The SRSWR Bootstrap was based on sample sizes of (m = n - 1), while the SRSWOR Bootstrap was based on sample sizes of (m = n/2).

47. The various variance estimation methods were evaluated against other variables from the "Quarterly Economic Activity Survey", as well as another survey using number-raised and ratio estimates, with similar findings. Therefore, it is recommended that the "best" variance estimation method is the SRSWOR Bootstrap for point-in-time estimates and the one component SRSWOR Bootstrap for movement estimates.

48. The second stage was to evaluate the "best" variance estimation method for point-in-time and movement of GREG estimates. The relative standard error percents, under the various variance estimation methods, based on approximately 50 and 100 replicate samples, for the point-in-time GREG estimates of sales from the "Quarterly Economic Activity Survey, March 2000" are presented in Table 3.

49. The estimated relative standard errors for the SRSWOR Bootstrap for the point-in-time GREG estimates were similar to those for the number-raised estimates. Therefore, it appears that the SRSWOR Bootstrap is also appropriate for point-in-time GREG estimates. At the time of this paper, the findings of the evaluation of the "best" variance estimation method for movement of GREG estimates was not available.

Table 3: Relative Standard Error Percents for Point-in-time GREG Estimates of Sales, Quarterly Economic Activity Survey, March 2000

Industry	Bootstrap SRSWOR	
	50	100
2	22.78	14.67
3	18.44	13.68
4	21.85	12.84
5	20.59	13.99
6	19.58	15.69
7	19.47	15.22
8	20.34	15.09
9	23.70	18.97
10	..	..
11	19.70	13.29
12	22.18	14.47
14	0.00	0.00
15	0.00	0.00
16	20.95	14.35
17	19.87	14.19
Total	**18.38**	**14.09**

4. Conclusion

50. The recommendation of this paper is that the computer system should use the Bootstrap (where the selection of replicate samples are by SRSWOR) to produce estimated variances of the generalised regression estimates. This method is acceptable when measured against the four characteristics required in the computing system and the underlying methodology mentioned at the start of the paper, with one exception. The exception is that it is not possible to conclude with sufficient power that the generalised regression point-in-time and movement variance estimators are unbiased.

51. The two alternative variance estimation methods are the Jacknife and Balanced Repeated Replication. The Jacknife is unsuitable because it requires many more replicate weights (i.e. the number of replicate weights needs to be at least equal to the sample size in each calibration class) for storage and processing. The Balanced Repeated Replication is unsuitable because of its simulation properties, specifically the efficiency of the variance estimator as measured by the root mean squared error.

52. The variance estimation methods used in computer systems of other statistical agencies do not have acceptable characteristics. The methodology supporting the Statistics Canada package, GES, is Taylor Series which does not meet the generic characteristic. Also the methodology supporting the Statistics Netherlands package, BASCULA, is Balanced Repeated Replication, which has already been shown to be unsuitable. The documentation accompanying BASCULA proposes an adjustment to Balanced Repeated Replication to improve its efficiency. This adjustment is called Balanced Repeated Replication with pseudo strata. However, the cost of improving efficiency will result in the method being unsuitable due to an increase in the number of replicate weights.

53. One variation of the Balanced Repeated Replication with pseudo strata is where the pseudo strata are partially, rather than completely balanced. The partial balancing arbitrarily reduces the number of replicate weights that are required by the complete balancing. However, the cost of partial balancing is a decrease in the efficiency of the variance estimator. Without a simulation study it is not possible to measure the efficiency of this variation to the Balanced Repeated Replication method.

54. Apart from the issues raised in this paper, there are a number of other issues that needed to be considered in relation to developing a computer system that will produce generalised regression estimates and measures of accuracy of these generalised regression estimates. These issues include developing methods to deal with surprise outliers, non-response, undercoverage of survey frames, imputation, winsorisation and non-convergence of the calibrated replicate weights.

5. Discussion Points for MAC Members

55. The questions for MAC members in relation to the developing of a methodology and a computer system that will produce generalised regression estimates and measures of accuracy of these generalised regression estimates are:

- Is there sufficient evidence to support the recommendation to adopt the Bootstrap (SRSWOR)? If not, what should be done to obtain this evidence?
- Is there sufficient theoretical and empirical evidence in the literature to support the hypothesis that the Bootstrap variance estimator is unbiased, or should the simulation study be extended?
- What is the minimum acceptable level of simulation error? (i.e. how many replicate weights should be used?)
- What should be done to determine this acceptable level?
- What brief guidance can be provided on the other issues listed in the conclusion?

6. Bibliography

Deville, J.-C. and Sarndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

Hidiroglou, M.A. and Sarndal, C.-E. (1998). Use of auxiliary information for two-phase sampling. *Survey Methodology*, 24, 11-20.

Hidiroglou, M.A., Estevao, V.M. and Arcaro, C. (2000). Generalised estimation systems and future enhancements. *Proceedings of the Second International Conference on Establishment Surveys (ICES II)*, 687-696.

Brewer, K.R.W., Gross, W.F. and Lee, G.F. (2000) PPN Sampling: The Australian Experience. *ISI Proceedings: Invited Papers, IASS Topics, Helsinki*, 155-163.

Rao, J.N.K. and Wu, C.F.J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83, 231-241.

Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, 18, 209-217.

Rust, K.F. and Rao, J.N.K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5, 283-310.

Sarndal, C.-E., Swensson, B. and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

Shao, J. and Tu, D. (1995) *The Jacknife and Bootstrap.* Springer, New York.

Singh, A.C. and Mohl, C.A. (1996). Understanding calibration estimators in survey sampling. *Survey Methodology*, 22, 107-115.

Wolter, K.M. (1985). *Introduction to Variance Estimation*. Springer-Verlag, New York.